Statistical upscaling of dynamic terrain state properties using a Symbolic Aggregate approXimation (SAX) approach

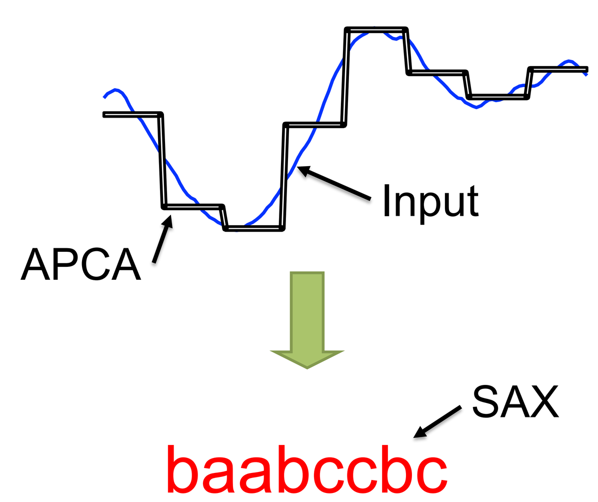
S. Frankenstein – ERDC/CRREL

Abstract

The current method for predicting terrain state properties such as soil moisture and temperature at the resolution and spatial scales needed for applications such as route planning and infrared sensor performance is computationally expensive due to the highly nonlinear physics involved. Areas of interest are not homogeneous and like pieces of ground over which representative terrain state information is calculated may, or may not be, contiguous. The complexity of the spatial and temporal variability of natural systems combined with the inherent nonlinear interactions in these systems makes it difficult to scale the problem. Not representing the heterogeneity, or representing the heterogeneity at a scale not commensurate with the model physics, can increase the uncertainty of the model predicted geophysical information. SAX (Symbolic Aggregate approXimation) of a data series has been successfully used to look for anomalies, changes and patterns in behavior. SAX is based on the concept of piecewise aggregate approximation in which the data is divided into a user determined number of segments, "w", having equal length, each of which is represented by its mean. The size of the symbolic alphabet, "k", is such that the distribution of segment means is Gaussian. The original time series is now represented by a "word" of length "w". Before assigning alphabets and word lengths, the different soil moisture and temperature time series are divided into subsections using Adaptive Piecewise Constant Approximation (APCA). A representative APCA curve is chosen then used to determine the SAX parameters from which the terrain state at other locations are predicted. No one has tried to use the SAX approach in this manner before. The all-season, dynamic 1-D state of the ground model FASST (Fast All-season Soil STrength) is used to provide the terrain response to predictive weather forcing. Being able to identify the spatial and temporal relationships in this fashion will help to account for the variability and enhance the fidelity (reduce the uncertainty) of the model predicted parameters.

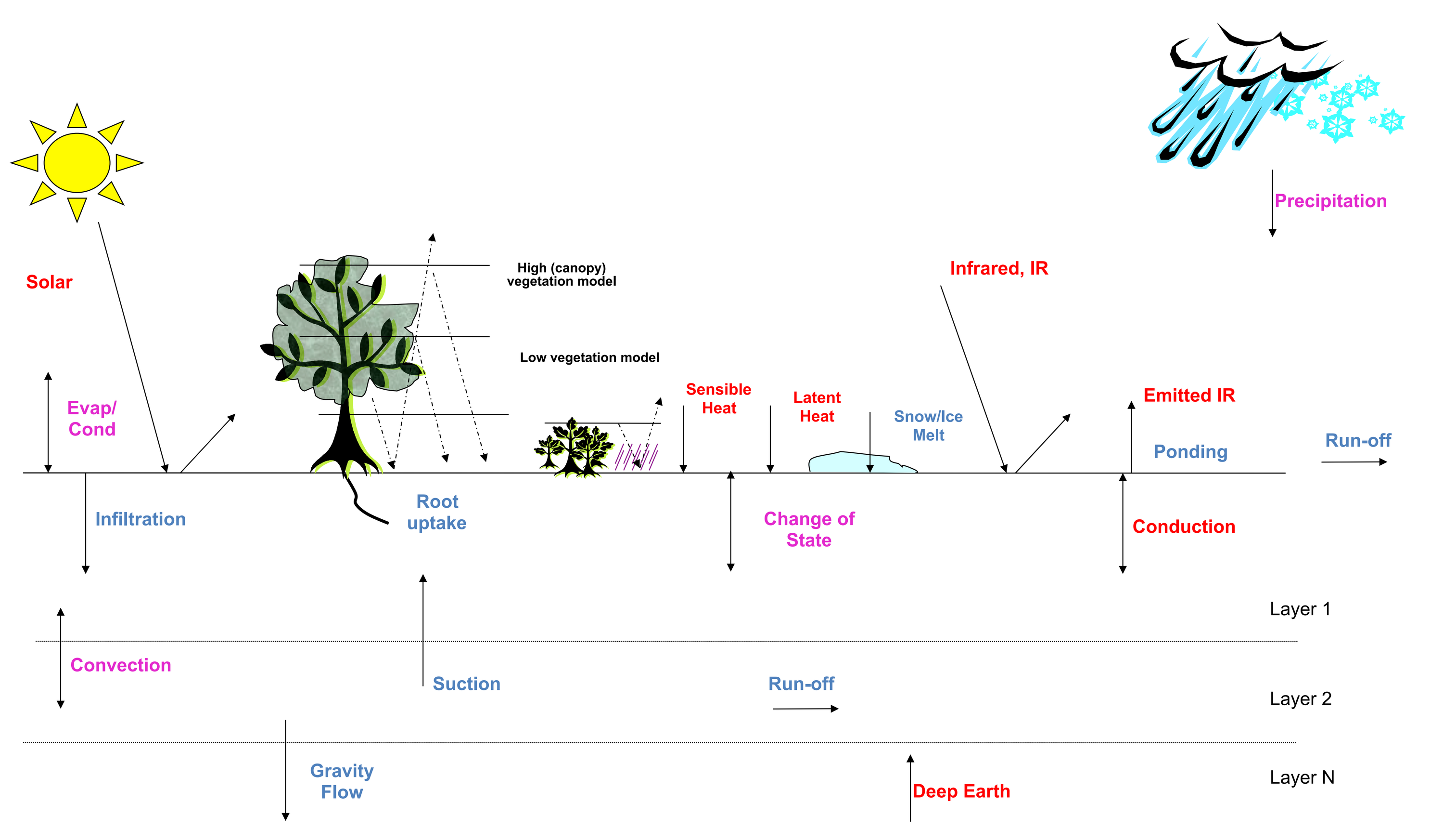
Symbolic Aggregate approXimation

- Developed by Eamonn Keogh (UCR; Keogh et al. 2001) & Jessica Lin (GMU; Lin et al. 2003)
- Original intent: Quick, efficient method to discover patterns/similarities in time series and shapes
- Create a APCA [Adaptive Piecewise Constant Approximation] (Chakrabarti et al. 2002)
  - Normalize data
  - Break data into linear segments of unequal lengths
- Determine length of "alphabet" (SAX)
- Look for patterns in SAX



To generate the soil moisture and temperature curves I used the Army's land surface model "FASST" – Fast All-season Soil Strength (Frankenstein and Koenig 2004a, 2004b; Frankenstein 2008). FASST is a full energy and mass balance model. Using forecast or climatological weather forcing, FASST outputs are being used to determine soil strength for mobility operations, road conditions for winter maintenance as well as sensor performance at various wavelengths. A schematic is shown below.

FASST Water & Energy Balance

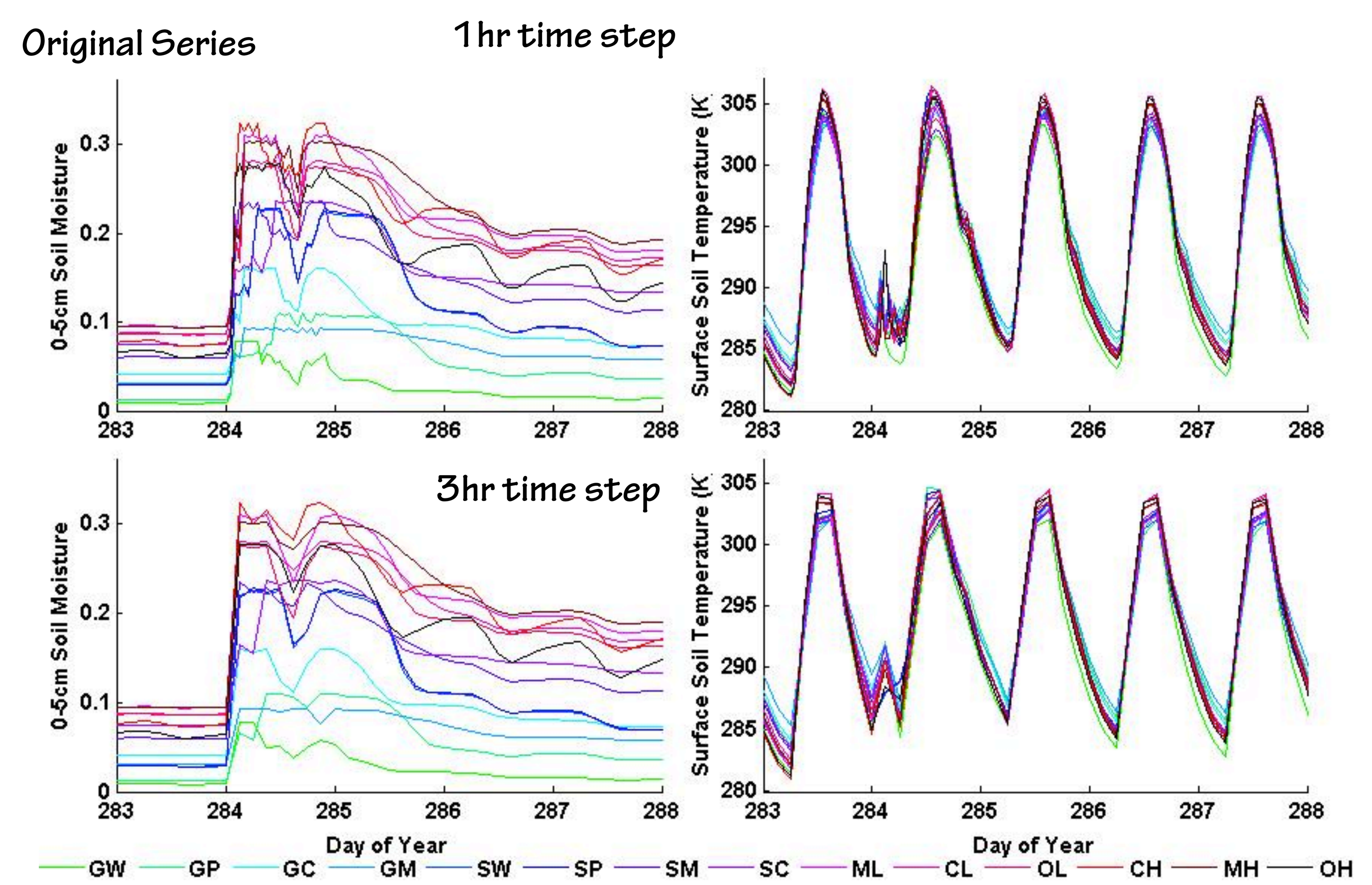


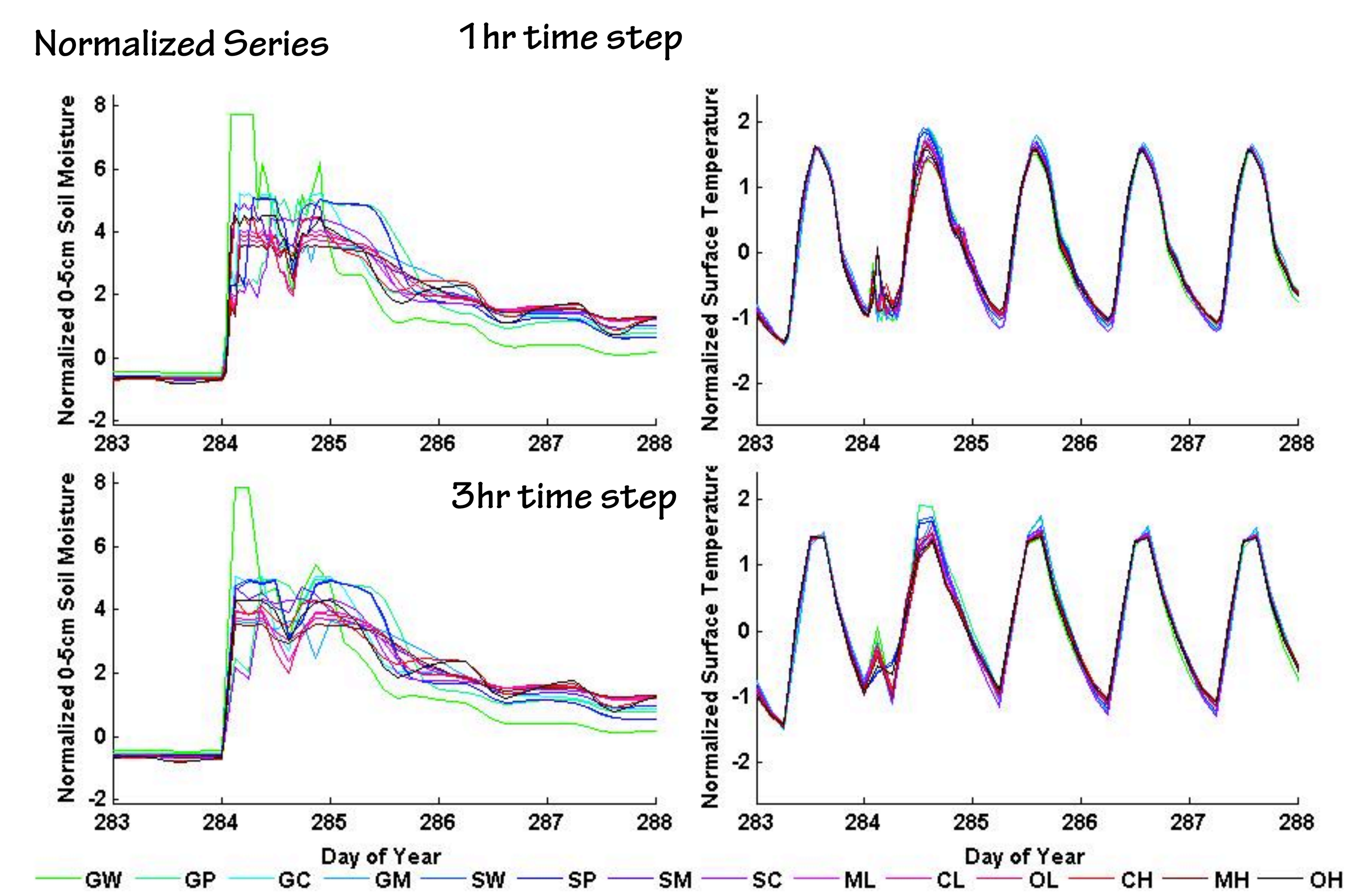
Weather data typically comes in either a 1 hour or 3 hour time step. Also, the Army gets its' forecast data from the Air Force Weather Agency (AFWA) which is based on WRF. Although they can go out further, typical forecasts are for 48 hours. Based on this information, I looked at the effects on the APCA and SAX series if

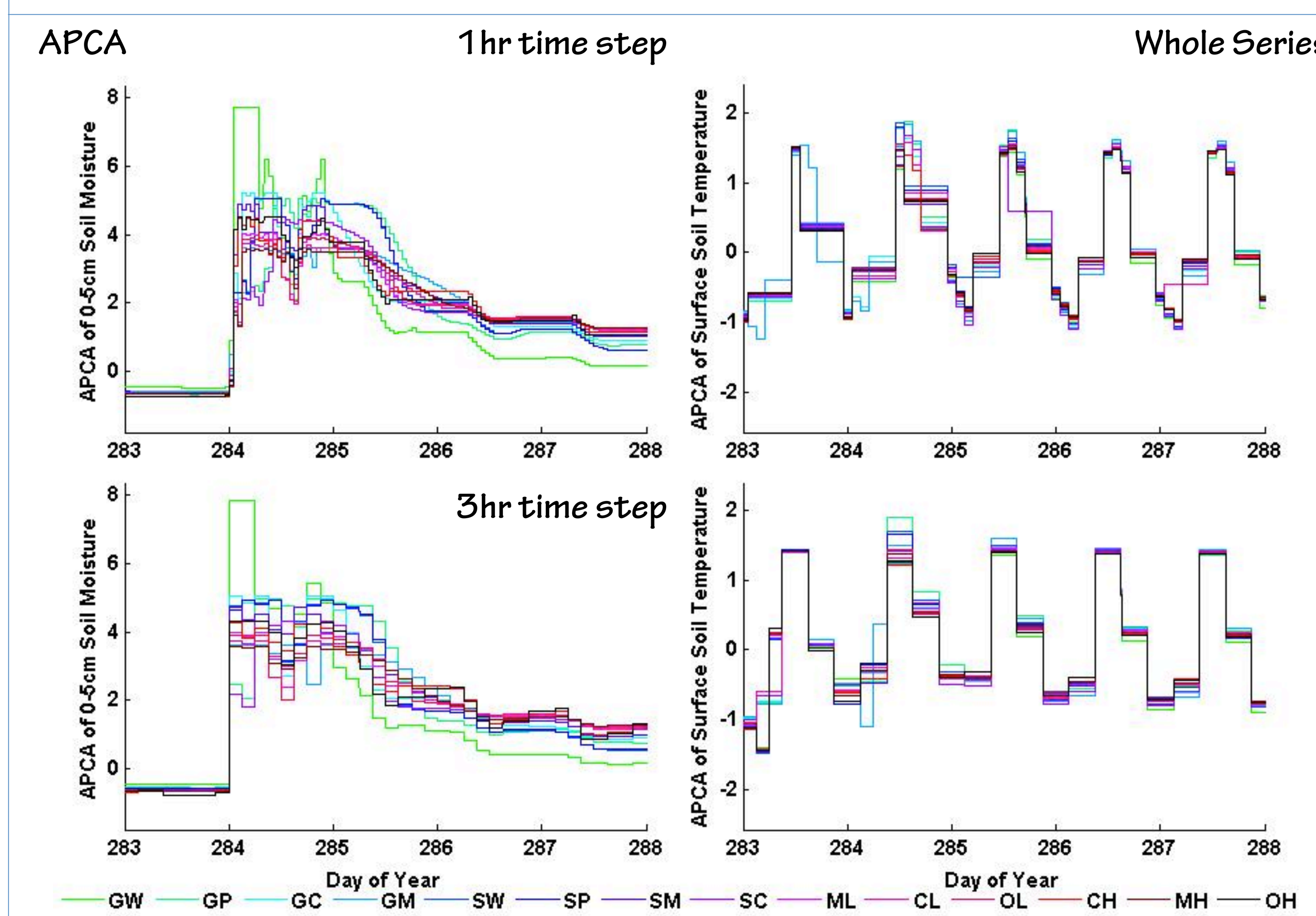
1. a 3 hour time step is used instead of a 1 hour step
2. the method is applied over the whole time series instead of multi 48 hour "forecasts"

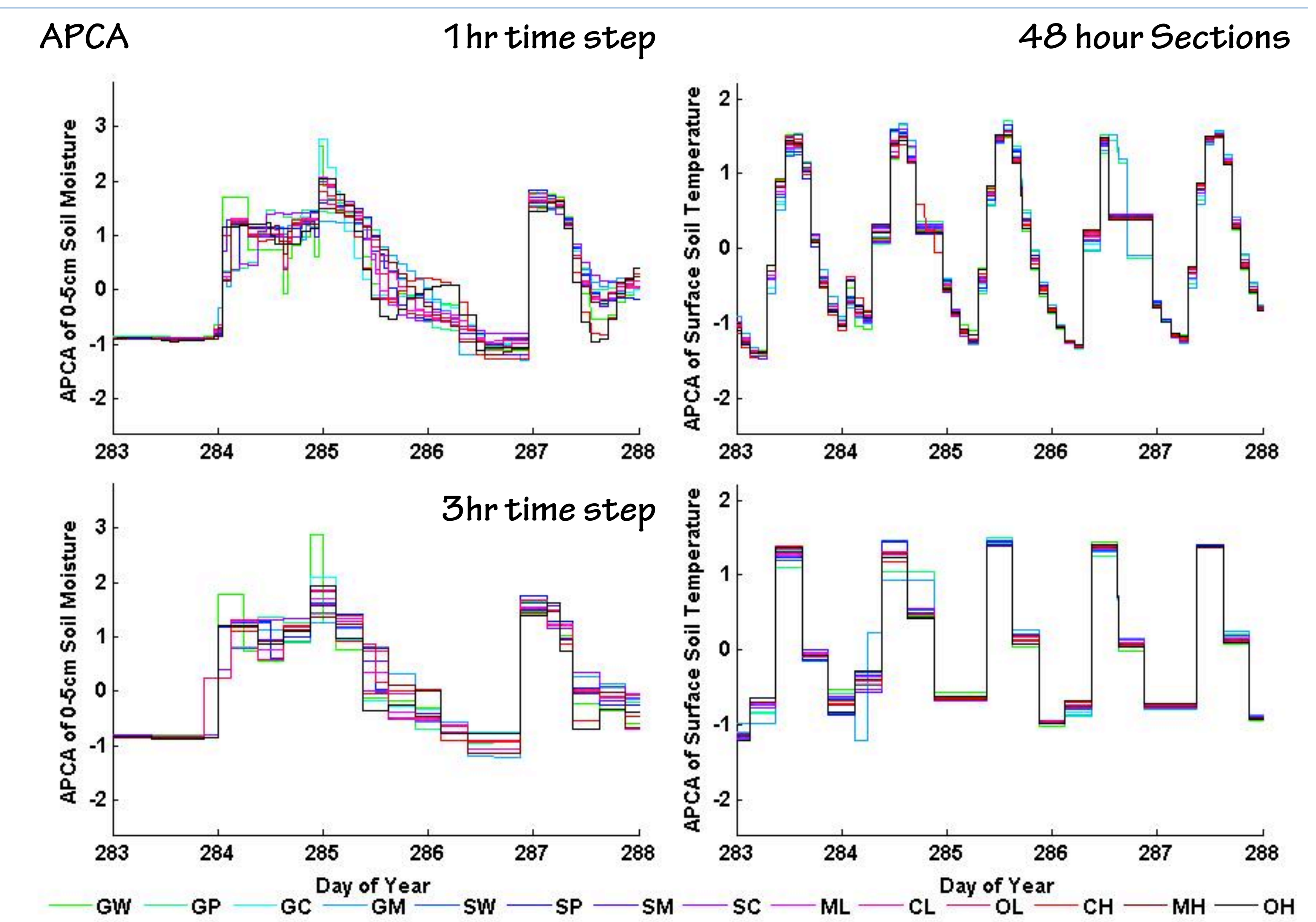
for both soil temperature and moisture. I also investigated the effect of soil type. FASST can model 18 different soils based on the engineering classification (Unified Soil Classification System) as well as bed rock, asphalt, concrete, permanent snow air (for tunnels and bridges) and inland water bodies. For this presentation, only the results for the14 main soils are shown. I generated the 1 hour weather forcing by taking a real measured day from Yuma, AZ and replicating it 30 times (DOY 259-299). On day 284 I inserted heavy rain. For the 3 hour data, I simply deleted the extra time steps.

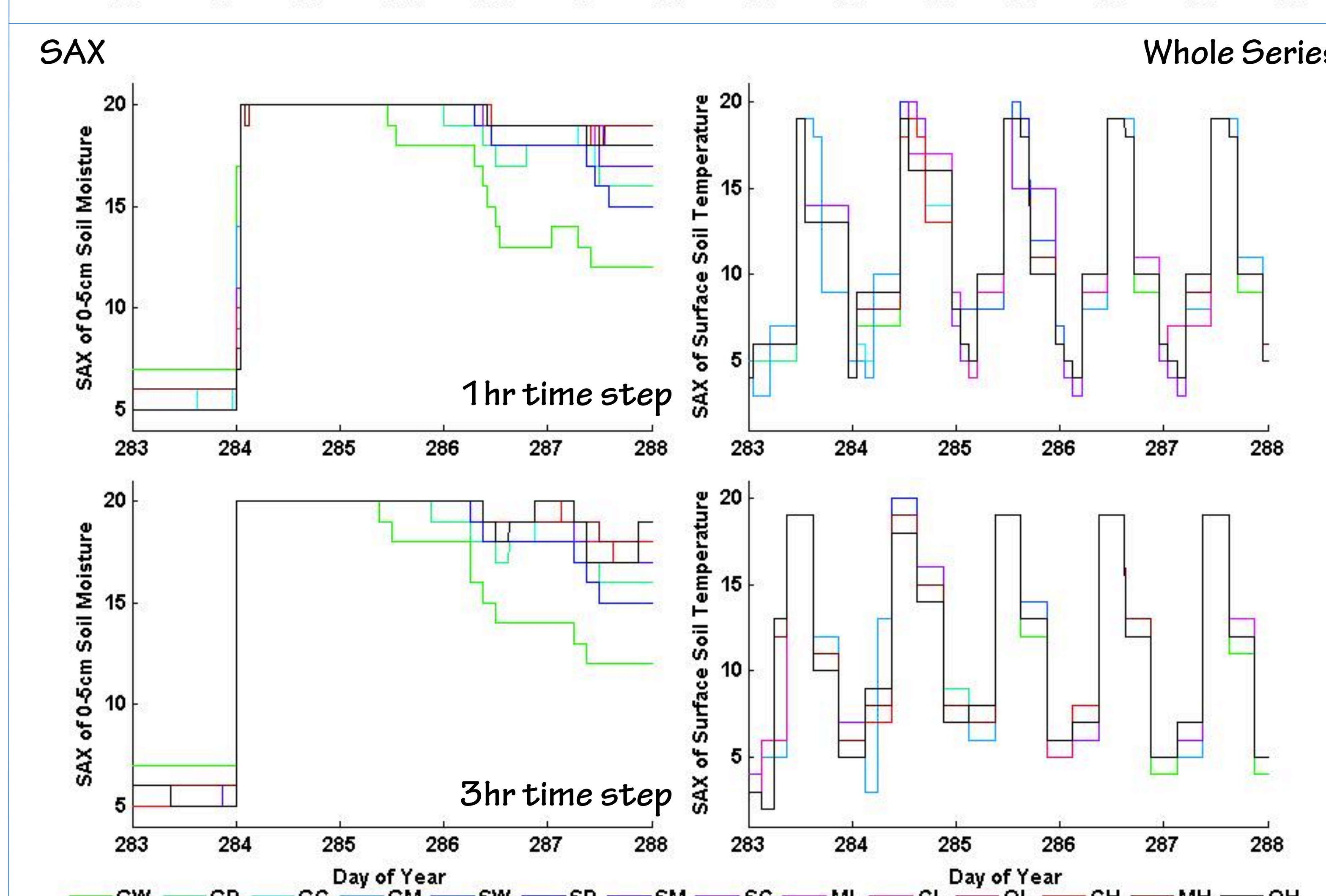
The "word length" is 320/16 for the 3 hour time step curves and 960/48 for the 1 hour curves. For all curves, the "alphabet" size is 20. In the figures, the numeric value is used instead of the "letter" to make viewing easier. Thus, 5 = "e", 10 = "j", 15 = "o" and 20 = "t". The word length is very important to capturing important features, especially the moisture curves. The weather forcing time step less so.

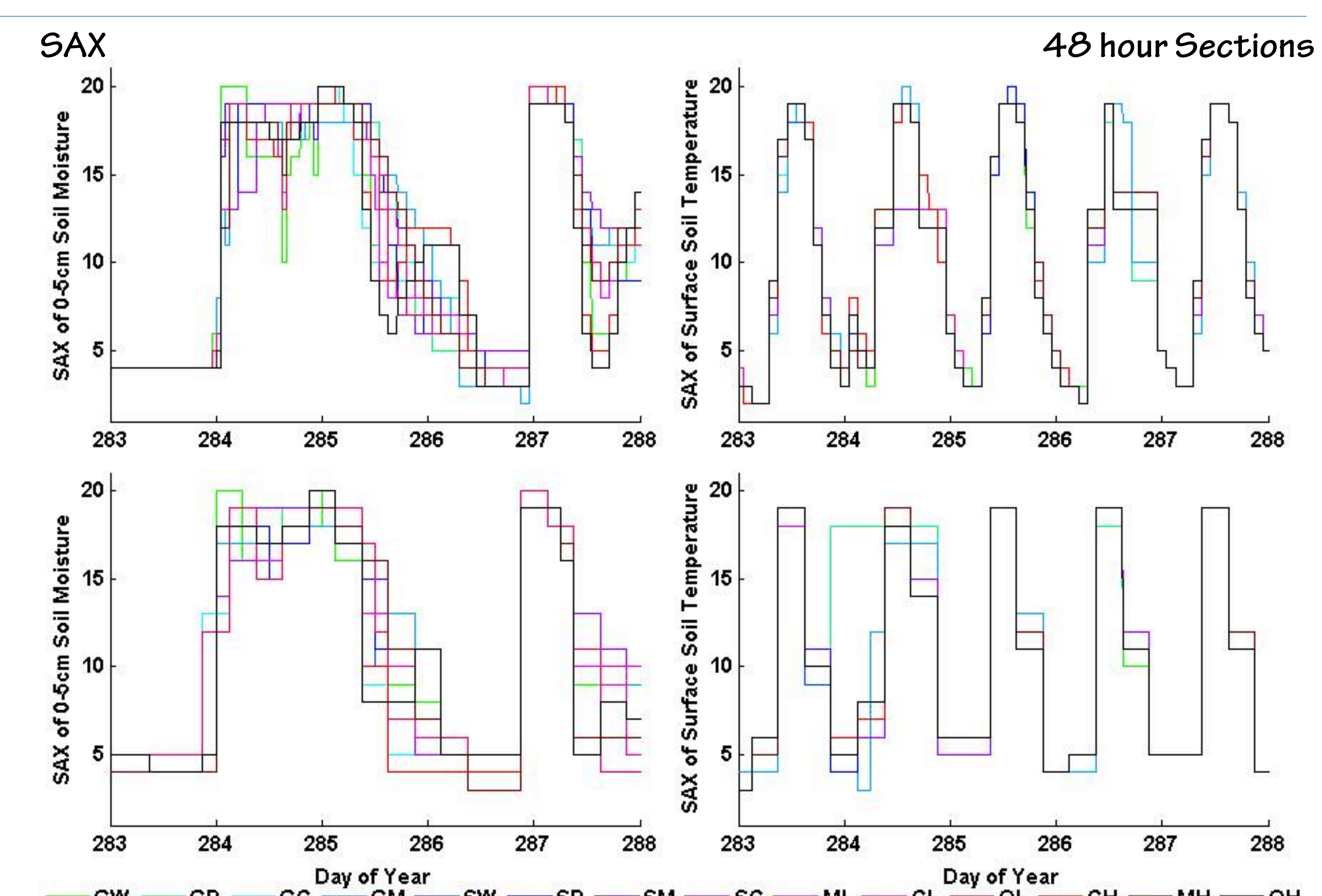












References

- Chakrabarti, K., E. Keogh, S. Mehrotra and M. Pazzani (2002) "Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases", ACM Trans. On Database Systems, 27(2), pp. 188-228.
- Frankenstein, S. (2008) "FASST Soil Moisture, Soil Temperature: Original vs. New", ERDC/CRREL Technical Report, TR-08-7, 38 pages.
- Frankenstein, S. and G. Koenig (2004b) "FASST Vegetation Models", ERDC/CRREL Technical Report, TR-04-25, 57 pages.

References (cont.)

- Frankenstein, S. and G. Koenig (2004b) Fast All-season Soil STrength (FASST)", ERDC/CRREL Special Report, SR-04-1, 107 pages.
- Keogh, E., K. Chakrabarti, M. Pazzani and S. Mehrotra (2001) "Dimensionality reduction for fast similarity search in large time series databases", J. of Knowledge and Information Systems, 3(3), pp. 263-286.
- Lin, J., E. Keogh, S. Lonardi and B. Chiu (2003) "A symbolic representation of time series, with implications for streaming algorithms", Proc. of 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA, pp. 2-11.